

Assessment of OvineSNP50 in Nigerian and Kenyan sheep populations

Ilori, B. M.^{1,2}, Rosen, B. D.³, Sonstegard, T. S.⁴, Bankole, O. M.¹, Durosaro, S. O.¹ and Hanotte, O.²

¹Department of Animal Breeding and Genetics, Federal University of Agriculture, Abeokuta, Ogun State Nigeria

²School of Life Sciences, University of Nottingham, NG7 2RD, Nottingham, UK.

³Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705, USA

Copyright resides with the authors in terms of the Creative Commons License 4.0.

(See <http://creativecommons.org/licenses/by/4.0/>).

Condition of use: The user may copy, distribute, transmit and adapt the work, but must recognize the authors and the Nigerian Journal of Biotechnology.

Abstract

Deciphering genomic information requires markers that are polymorphic and sufficient enough to capture its vast array of genetic data. Polymorphic loci can differ greatly between breeds of the same species and the exclusion of the Nigerian and some African sheep breeds during the development of the OvineSNP50 chip necessitated the validation of SNPs included on the chip to allow for genomic applications of the excluded breeds. A total of sixty sheep samples were genotyped [10 each of the Balami, Uda, West African Dwarf and Yankasa from Nigeria, Dorper and Red Maasai from Kenya (East Africa)] using the Ovine 50k Illumina SNP bead chip. Results revealed that 33,994 SNPs (97.47%) of the called 34,876 SNPs were validated for downstream analysis. Mean heterozygosity values of 0.154 and 0.153 were obtained for polymorphic SNPs on sex and autosomal chromosomes respectively, while the values of 0.662 and 0.054 were obtained on the sex and autosomal chromosomes respectively for the mean identity-by-state (IBS). Six and three individuals violated the per ID and identity-by-state (IBS) thresholds, respectively. It was observed that the Ovine 50k Illumina SNP bead chip was informative in the Nigerian and East African sheep that were studied, and should be useful in examining the underlying genetic variation.

Keywords: Sheep, OvineSNP50, genome-wide, call rate, minor allele frequency

***Author for Correspondence:** iloribm@funaab.edu.ng

Introduction

Genetic studies in Nigerian sheep has been extensive and repeatedly on phenotypic characteristics (Yakubu and Ibrahim, 2011; Adejoro and Salako, 2012; Yunusa et al., 2013) and some genetic diversity studies using microsatellite (Agaviezor et al., 2012a) and mitochondrial markers (Agaviezor et al., 2012b). The genetic basis which underpins this diversity and the consequence of selection on the genetic variation present within the sheep breeds have not been assayed on a genome-wide basis. To date, no studies have reported levels of Nigerian sheep genetic diversity using autosomal

variation on a genome-wide scale. The genomic abundance and amenability to cost effective high throughput genotyping has indicated that single nucleotide polymorphisms (SNP) are now the most widely used class of genetic markers (Kijas et al., 2009). The OvineSNP50 developed by Illumina in collaboration with the International Sheep Genomics Consortium in 2009 was designed with genotype of 54,000 polymorphisms with a mean genomic distance of 51 kb across the ovine genome (International Sheep Genomics Consortium et al., 2010; Kijas et al., 2012). The SNP panel involves 33,000 SNPs of six samples in six different breeds discovered

by Roche 454 whole-genome shotgun sequencing. Fifteen thousand (15,000) SNPs identified reduced representational sequencing on 60 individuals from 15 breeds of sheep and 500 SNPs detected using Bacterial Artificial Chromosome and Sanger sequencing of nine individuals from different breeds were incorporated into the Ovine SNP50 chip (Kijas et al., 2012).

SNP genotyping is a powerful tools and as molecular markers which are based on DNA sequence polymorphisms and which must be sufficient and spaced evenly throughout the genome for it to be informative for population based studies (Kijas et al., 2009). It is useful for the analysis of genetic biodiversity, population structure, linkage disequilibrium, mapping quantitative trait loci and until recent past genomic selection. Indeed, DNA sequences determine the diversity of organisms, and therefore, the techniques used to evaluate DNA polymorphisms directly measure the genetic diversity. With limited ovine genomic sequence available, the International Sheep Genomics Consortium's immediate objective has been to sequence the ovine genome so as to identify SNPs in order to produce a 50k SNP chip. The Ovine SNP50 Bead Chip is a comprehensive genome-wide genotyping array for the ovine genome, providing superior power to interrogate genetic variation across many breeds. Data generated from the Ovine SNP50 Bead Chip have been used in several ways, including the development of Molecular Breeding Values, or the identification of genomic regions which explain either or all of the variation in monogenic traits or a large amount of variation in polygenic traits. There are already a number of publications that report the identification of the gene causing various monogenic disorders such as Arthrogryposis, Achondroplasia and Progressive Muscular Dystrophy through the selective genotyping of case and control animals using the Ovine SNP50 Bead Chip (Christensen et al., 2013).

The four main breeds of sheeps in Nigeria namely Balami, Uda, Yankasa and West African Dwarf (Adu and Ngere, 1979) with an additional two breeds from Kenya; Dorper and Red Maasai were included in this study to

recognize the possible effect of ascertainment bias. This is possible as the chip was developed from a limited number of breeds and samples which are not true representative of the polymorphism and frequency of alleles in the genome of members of this species (Nielsen and Signorovitch, 2003).

Knowledge of the genetic relationship among neighbouring sheep populations is crucial for conservation efforts (Mukhongo et al., 2014) while genetic diversity serves as a way for populations to adapt to changing environments. With more variations, it is more likely that some individuals in a population will possess alleles that are suited for the environment (Fu et al., 2017). The conservation of farm animal genetic resources is important for coping with future breeding needs and for facilitating the sustainable use of marginal areas.

With the exclusion of the Nigerian sheep breeds during the development of the ovine 50k SNP bead chip, ascertainment bias of SNPs in the panel and the Nigerian Sheep Genome study which is lacking, it is needed to evaluate Nigerian and neighbouring sheep genetic resource with the Ovine 50k SNP bead chip to determine whether they are polymorphic for sufficient number of SNPs genotype in the bead chip, its applicability and its suitability to characterise this resource which could pave way for their conservation and better breeding management.

Materials & Methods

Experimental Population

Blood samples were collected from four indigenous sheep breeds (Balami, Uda, West African Dwarf and Yankasa) and two East African sheep breeds (Dorper and Red Maasai). Ten animals above tow years of age per breed were sampled making a total of sixty animals. For the Nigerian indigenous breeds, samples were gotten from College of Animal Science and Livestock Production (COLANIM) farm at the Federal University of Agriculture, Abeokuta (FUNAAB) and Kuto market also in Abeokuta, while the East African samples were gotten from the International Livestock Research Institute (ILRI) in Kenya.

Blood Sample Collection, DNA Extraction and genotyping

Blood samples were collected from the animals into heparinised tubes from the jugular vein and stored at -2°C using EDTA (ethylenediamine-tetra-acetic acid) as anticoagulants. Approximately, 2ml of blood was collected from each of the animals into EDTA bottles, then transferred to the Animal Breeding and Genetics laboratory (FUNAAB) for analysis. DNA was extracted using NORGEN Genomic DNA isolation kit (NorgenBiotek Corporation, Ontario, Canada) according to the manufacturer's protocols. Concentration and purity of the extracted DNA were ascertained using gel electrophoresis and NanoDrop. Genotyping of DNA samples was outsourced to a private sequencing company using Illumina Infinium array (GeneSeek, Lincoln, NE, USA) consisting of nearly 50,000 SNPs derived mainly from Ovine breeds according to Illumina standard protocols. GenomeStudio Software v1.0 (Genotyping Module Illumina) was used for genotype calling from SNP intensity data and to ensure stringency of quality control parameters.

SNP Quality Control and downstream analysis

The SNP data generated were subjected

to quality control using the following criteria to determine the number of polymorphic loci: Minor Allele Frequency (MAF) - 1%, SNP Call Rate - 90%, Per Individual Call rate or sample call rate - 90%, Identity-by-state (IBS) - 90%. The quality control measures were ensured using the check.marker function of GenABEL package implemented in R software (Aulchenko et al., 2007). MAF and within breed genetic variability (Observed and Expected Heterozygosity) was estimated using descriptive.marker and per.id summary function of GenABEL Package implemented in R software.

Results

SNP Genotyping

From the 60 samples outsourced for genotyping, Balami and Yankassa samples did not amplify and therefore were excluded from the analysis. A total of 34,876 SNPs were mapped including: 865 on the X chromosome and 34,011 on the 26 autosomal chromosomes. Both sets were analysed separately. The distribution of genotyped markers on all chromosomes is presented in Table 1. SNP genotyped ranged from 421 on chromosome 24 to 3815 on chromosome 1.

Table 1: Distribution of markers on autosomal and sex chromosomes

Chromosome	Distribution of markers		
	Before quality check	After quality check	Number removed
1	3815	3732	83
2	3766	3686	80
3	3271	3183	88
4	1801	1756	45
5	1529	1499	30
6	1867	1821	46
7	1510	1479	31
8	1441	1409	32
9	1391	1364	27
10	1221	1183	38
11	717	697	20
12	1102	1068	34
13	1063	1034	29
14	673	645	28
15	1050	1022	28
16	955	929	26

17	865	838	27
18	914	888	26
19	766	746	20
20	692	681	11
21	520	507	13
22	726	707	19
23	694	680	14
24	421	411	10
25	643	630	13
26	598	581	17
X	865	818	47

SNP Quality Control

The result of the sample-based quality control (Table 2) showed that out of the 865 SNPs on the sex chromosome, 1 marker violated the call rate threshold of 90% while 46 markers were removed for violating the minor allele frequency threshold of 1%. Six individuals did not meet the 90% threshold of per individual calling rate (Per. ID) and another three individuals violated the identity-by-state (IBS) 90% threshold, with the mean IBS being 0.662 and the mean heterozygosity of 0.154. Nine individuals and 47 markers were removed for

violating quality control remaining 818 markers on sex chromosome.

On the autosomes, 33 markers were removed due to low call rate and 802 markers were removed for violating minor allele frequency. Also, six individuals violated the 90% threshold of per individual calling rate (Per I.D) and another three violated the identity-by-state (IBS) 90% threshold, with the mean IBS being 0.054 and the mean heterozygosity at 0.153. A total of 9 individuals, which were the same as for sex chromosome and 835 markers were removed for violating quality control remaining 31 individuals and 33,176 markers.

Table 2: Quality control of SNPs

Chromosome	Number of SNPs	SNPs Call Rate <90%	SNPs MAF <1%	Per I.D	IBS	Mean IBS	Mean Heterozygosity	Total SNPs removed	SNPs validated
X	865	1	46	6	3	0.662	0.154	47	818
Autosome	34,011	33	802	6	3	0.054	0.153	835	33,176

Distribution of minor allele frequency of markers within the populations

MAF distribution across the populations on the Autosomal chromosomes

It was observed that with the MAF of less or equal to 1%, the proportion of SNPs ranged from 0.073 in Dorper to 0.216 in WAD (Table 3). It was also observed that with the MAF of greater than 1% and less or equal to 5%, only WAD has a

proportion of 0.094. Uda has the highest proportion of 0.179 compared to WAD's 0.088 with the MAF of greater than 5% and less or equal to 10%. With the MAF of greater than 10% and less or equal to 20%, SNP proportion ranged from 0.131 in Red Maasai to 0.204 in Dorper. Finally, SNP proportion ranged from 0.437 in WAD to 0.631 in Dorper with the MAF of greater than 20%.

Table 3: Minor allele frequency distribution across the populations' autosomal chromosomes.

MAF	Dorper		Red Maasai		Uda		WAD	
	SNP	Prop	SNP	Prop	SNP	Prop	SNP	Prop
0.01 < X ≤ 0.05	2409	0.073	5589	0.168	5721	0.172	7177	0.216
0.05 < X ≤ 0.1	0	0	0	0	0	0	3112	0.094
0.1 < X ≤ 0.2	3067	0.092	4338	0.131	5947	0.179	2928	0.088
> 0.2	6761	0.204	4354	0.131	5997	0.181	5456	0.164
Total	20939	0.631	18895	0.570	15511	0.468	14503	0.437
	33176		33176		33176		33176	

Prop = proportion

MAF distribution across the populations on the sex chromosome

On the sex chromosome, It was also observed that with the MAF of less or equal to 1%, the proportion of SNPs ranged from 0.164 in Dorper to 0.248 in Uda (Table 4). It was also discovered that with the MAF of greater than 1% and less or equal to 5%, only WAD has a

proportion of 0.104. Red Maasai has the highest proportion of 0.149 compared to Dorper's 0.044 with the MAF greater than 5% and less or equal to 10%. With the MAF of greater than 10% and less or equal to 20%, SNP proportion range from 0.123 in Red Maasai to 0.233 in Dorper. Finally, SNP proportion ranged from 0.423 in WAD to 0.559 in Dorper with the MAF of greater than 20%.

Table 4: Minor allele frequency distribution across the four populations' sex chromosome.

MAF	Dorper		Red Maasai		Uda		WAD	
	SNP	Prop	SNP	Prop	SNP	Prop	SNP	Prop
X ≤ 0.01	134	0.164	164	0.200	203	0.248	200	0.244
0.01 < X ≤ 0.05	0	0	0	0	0	0	85	0.104
0.05 < X ≤ 0.1	36	0.044	122	0.149	52	0.064	83	0.101
0.1 < X ≤ 0.2	191	0.233	101	0.123	174	0.213	104	0.127
> 0.2	457	0.559	431	0.527	389	0.476	346	0.423
Total	818		818		818		818	

Prop = proportion

Discussion

The OvineSNP50 which is a microarray-based system was designed to determine the genotype of approximately 54 000 single nucleotide polymorphisms spaced evenly across the ovine genome, yet excluding the Nigerian and other African sheep populations (Sandenbergh et al., 2016). Sample-based filtering relates to the call rate, which is the fraction of called SNPs per sample over the total number of SNPs in the dataset and can be influenced by factors such as DNA quality and concentration (Anderson et al., 2010). The call rate of 99.90% obtained in this study is comparable with call rates of above 99.5% (Kijas et al., 2013) and 99.9% (Tosser-Klopp et al., 2014) reported in related studies. It also compares favourably with the call rate of 99.6% found in the South African Angora goats (Lashmar et al., 2015). This indicates the high quality and informativeness of the Nigerian sheep populations' DNA.

The successful application of SNP arrays depends largely on the degree of polymorphism in the various breeds within each species (Fan et al., 2010). After filtering out SNPs with low minor allele frequency (MAF), which is the frequency at which the second most common allele occurs in a population, 97.57% proved to be polymorphic for these sheep breeds which is comparable to the number of polymorphic loci found in South African populations (81.16–86.85%; Sandenbergh et al., 2016)) and goat breeds (Jinlan: 45,648, Skopelos: 50,908) that were not included in the SNP discovery process (Tosser-Klopp et al., 2014). The number of polymorphic loci compares favourably with values of 28,869 and 35,084 SNPs obtained in African N'Dama and Sheko breeds, respectively, during Illumina's bovine SNP50 content validation study (Matukumalli et al., 2009) and the result of this study indicates a high level of polymorphism which is due to the informativeness of the African breeds genotyped.

The Per. Individual call rate (*Per. I.D*) refers to a quality control measure used in pinpointing the precision of downstream data analyses in genomic prediction and the *Per. I.D* call rate of the population at 85%, pinpoint towards an informative population. When two or more individuals have identical nucleotide

sequences in a DNA segment, they are said to be identical by state (IBS). For any given pair of individuals with genotype information, identity-by-state (IBS) can be observed at a given locus with three possible outcomes: the individuals have two different alleles or they share one or two alleles in common. Two individuals who share 1 or 2 alleles IBS at a given locus may have inherited the shared allele(s) from a recent common ancestor, in which these allele(s) are identical-by-descent (IBD) (Lashmar et al., 2015). After filtering out individuals that are identical by state, 92.5% of the individual population are validated for further analysis.

Heterozygosity is a condition where an organism has two different alleles of a given gene, which points to genetic variability within the population. The mean heterozygosity which is the average proportion of heterozygous SNPs in this study is high and indicates high level of genetic variation within the population with low level of inbreeding. The heterozygosity value obtained in the present study compares favourably with that of South African sheep (0.33 to 0.35) (Sandenbergh et al., 2016) and 0.365 obtained for Angora goats (Lashmar et al., 2015).

The MAF distribution patterns of all breeds were similar to those of South African, European and American breeds (Kijas et al., 2012; Sandenbergh et al., 2016) with most loci exhibiting MAFs of at least 20% or greater. However, West African Dwarf sheep exhibited lowest level of MAF distribution compared to other sheep breeds used in this study. The result might be explained by the low number of individuals and breeds of African origin used in the development of this chip panel resulting in SNP ascertainment bias and possibility of rare alleles in the WAD and other sheep populations used in the study. From our whole-genome genotyping result, it can be concluded that Nigerian and Kenyan sheep populations are polymorphic for majority of the SNPs genotyped on the SNP chip despite the fact that none of the breeds was included in its development. The WAD and Uda sheep breeds of Nigerian relatively had low MAF, heterozygosity and polymorphic marker when compared with the Kenya sheep breeds used in this study. This may most likely be

as a result of SNP ascertainment bias and the fact that these sheep breeds have not been improved for any known traits. The improper management might have resulted in inbreeding within the herd and resultant increase in non-polymorphic loci. Although effective population size and rate of evolution differ between autosome and X chromosomes, this marker allow for access to large number of genomic information although with ascertainment bias in Nigerian sheep populations which could be ameliorated using linkage pruning technique. It is recommended that the marker should be used in all Nigerian sheep populations for their characterization, population structure and marker assisted selection for trait of economic importance.

REFERENCES

- Adejoro, F. A. and Salako, A. E. (2012). Morphobiometric Characterization Of The Balami Sheep of Nigeria. *Afr. J. of Gen. Agric.* 8(4), December 31, 2012
- Agaviezor, B., Peters, S. O., Adefenwa, M. A., Yakubu, A., Adebambo, O. A., Ozoje, M. O., Ikeobi, C. O. N., Wheto, M., Ajayi, O. O., Amusan, S. A., Ekundayo, O., Sanni, T. M., Okpeku, M., Onasanya, G. O., Donato, M. D., Ilori, B. M., Kizilkaya, K., Imumorin, I. G. (2012a). Morphological and microsatellite DNA diversity of Nigerian indigenous sheep. *J. Anim. Sci. Biotechnol.* 3(38): 2–16
- Agaviezor, B. O., Adefenwa, M. A., Peters, S. O., Yakubu, A., Adebambo, O. A., Ozoje, M. O., Ikeobi, C. O. N., Ilori, B. M., Wheto, M., Ajayi, O. O., Amusan, S. A., Okpeku, M., Onasanya, G. O., Donato M. D., Imumorin, I. G. (2012b). Genetic diversity analysis of the mitochondrial D-loop of Nigerian indigenous sheep. *Ani. Genet. Res.* 50:13-20.
- Adu, I. F. and Ngere, L. O. (1979). The indigenous Sheep of Nigeria. *World Rev. of Anim. Prod.* 15(3): 51-62.
- Anderson, C. A., Pettersson, F. H., Clarke, G. M., Cardon, L. R., Morris, A. P. and Zondervan, K. T. (2010). Data quality control in genetic case-control association studies. *Nat Protoc.* 5(9): 1564-73. doi: 10.1038/nprot.2010.116. Epub 2010 Aug 26
- Aulchenko, Y. S., Ripke, S., Isaacs, A., van Duijn, C. M. (2007). GenABEL: an R library for genome-wide association analysis. *Bioinformatics* 23: 1294-6.
- Christensen, K. A., Brunelli, J. P., Lambert, M. J., DeKoning, J., Phillips, R. B. and Thorgaard, G. H. (2013). Identification of single nucleotide polymorphisms from the transcriptome of an organism with a whole genome duplication. *BMC Bioinformatics.* 14: 325
- Fan, B., Zhi-Qiang, D., Danielle, M., Gorbach, M. and Rothschild, F. (2010). Development and application of high-density SNP arrays in genomic studies of domestic animals. *Asian-Australas J. Anim. Sci.* 23: 833-847.
- Kijas J. W., Townley D., Dalrymple B. P., Heaton M. P., Maddox J. F., McGrath A., Wilson P., Ingersoll R. G., McCulloch R., McWilliam S., Tang D., McEwan J., Cockett N., Oddy V. H., Nicholas F. W. & Raadsma H. (2009). A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS One* 4, e4668.
- Kijas, J. W., Lenstra, J. A., Hayes, B., Boitard, S., Porto Neto, L. R., San Cristobal, M., Servin, B., McCulloch, R., Whan, V., Gietzen, K., Paiva, S., Barendse, W., Ciani, E., Raadsma, H., McEwan, J., Dalrymple, B. and other members of the International Sheep Genomics Consortium. (2012). Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS One* DOI: 10.1371/journal.plosbiology.1001258
- Kijas, J. W., Ortiz, J. S., McCulloch, R., James, A., Brice, B., Swain, B., Tosser-Klopp, G. and International GoatGenome Consortium. (2013). Genetic diversity and investigation of polledness in divergent goat populations using 52 088 SNPs. *Anim. Genet.* 44(3): 325-35. doi: 10.1111/age.12011
- Lashmar, S. F., Visser, C. and van Marle-Köster, E. (2015). Validation of the 50k Illumina goat SNP chip in the South African Angora goat. *S. Afr. J. Anim. Sci.* 45(1)



Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., Taylor, J. F., Allan, M. F., Heaton, M. P., O'Connell, J., Moore, S. S., Smith, T. P., Sonstegard, T. S. and Van Tassell, C. P. (2009). Development and characterization of a high density SNP genotyping assay for cattle. *PLoS One*. 4 (4) : e 5 3 5 0 . d o i : 10.1371/journal.pone.0005350.

Mukhongo, D. M., Mwai, O., Tapio, M. and Muigai, A. (2014). Genetic Diversity and Population Structure of the Indigenous Sheep in Kenya Based on Microsatellite Analysis: Implications for their Conservation. *Livest Sci*. 5: 65-78

Nielsen, R., Signorovitch, J. (2003). Correcting for ascertainment biases when analyzing SNP data: applications to the estimation of linkage disequilibrium. *Theor. Popul. Biol*. 63: 245–255.

Sandenbergh, L., Cloete, S. W. P., Roodt-Wilding, R., Snyman, M. A. and Bester-van der Merwe, A. E. (2016). Evaluation of the OvineSNP50 chip for use in four South African sheep breeds. *S. Afr. J. Anim. Sci.* 46 (1) : 89 - 93 , doi:10.4314/sajas.v46i1.11.

Tosser-Klopp, G., Bardou, P., Bouchez, O., Cabau, C., Crooijmans, R., Dong, Y., Donnadieu-Tonon, C., Eggen, A., Heuven, H.C.M., Jamli, S., Jiken, A.J., Klopp, C., Lawley, C.T., McEwan, J., Martin, P., Moreno, C.R., Mulsant, P., Nabihoudine, I., Pailhoux, E., Palhiere, E., Rupp, R., Sarry, J., Sayre, B.L., Tircazes, A., Wang, J., Wang, W., Zhang, W.G. and the International Goat Genome Consortium. (2014). Design and Characterization of a 52K SNP Chip for Goats. *PLoS One* 9 2014 - ISSN 1932-6203

E, G. X., Huang, Y. F., Zhao, Y. J. He, N. (2017). Genetic Diversity of three Chinese Native Sheep Breeds. *Russ. J. of Genet*. 53: 118 <https://doi.org/10.1134/S1022795417010045>

Yunusa, A. J., Salako, A. E. and Oladejo, O. A. (2013). Morphometric characterization of Nigerian indigenous sheep using multifactorial discriminant analysis. *Int. J. Biodivers. Conserv*. 5(10): 661-665.

Yakubu, A. and Ibrahim, I. A. (2011). Multivariate analysis of morphostructural characteristics in Nigerian indigenous sheep. *Italian J. Anim. Sci*. 10(2): 83-86.